Context-Aware Heterogeneous Text Representation for Vocational Internship Placement Recommendation

Asrul Helmandi, Andysah Putera Utama Siahaan

Abstract

Vocational internship allocation serves as a vital bridge connecting student competency development with industrial requirements in business and corporate sectors. Nonetheless, conventional allocation workflows encounter substantial bottlenecks caused by heterogeneous text disparities and semantic gaps between fragmented student profiles (majors, core skills, vocational subjects, and internship interests) and free-form narrative industry vacancy descriptions. Furthermore, the absence of prior vocational certification among 11th-grade students induces severe cold-start vulnerability. This paper proposes a Context-Aware Heterogeneous Text Representation framework that integrates multi-feature semantic embedding spaces with an operational context gating mechanism (major relevance and placement calendar periods). Validated on real-world institutional datasets from SMK Negeri 8 Medan (653 eleventh-grade students and 103 partner industry entities in 2025 across Culinary Arts, Fashion Design, Beauty Therapy, and Hospitality), empirical results demonstrate that the proposed framework achieves 98.2% Precision@1 and an NDCG@3 of 0.964, outperforming monolithic TF-IDF baselines (74.5% Precision@1) and exact keyword matching (61.8% Precision@1). The operational gating mechanism eliminates domain violation rates to 0.0%, establishing an accountable, automated, and precision-driven methodology for vocational internship placement.

Keywords: Vocational Education, Internship Placement, Heterogeneous Text Representation, Semantic Alignment, Recommender Systems, Industry Matching

Asrul Helmandi[1]
[1]Magister Teknologi Informasi, Universitas Pembangunan Panca Budi, Indonesia
e-mail: asrulhelmandiacc@gmail.com [1]

Andysah Putera Utama Siahaan[2]
[2]Magister Teknologi Informasi, Universitas Pembangunan Panca Budi, Indonesia
e-mail: andiesiahaan@gmail.com [2]
2nd International Conference on Islamic Community Studies (ICICS)
Theme: History of Malay Civilisation and Islamic Human Capacity and Halal Hub in the Globalization Era
https://proceeding.pancabudi.ac.id/index.php/ICIE/index

Introduction

Vocational education plays a strategic role in preparing an adaptive and skilled workforce whose competencies are aligned with the evolving needs of the Business and Industrial Sectors (DUDI) [8]. One of the essential components of the vocational education curriculum that serves as a bridge between formal schooling and the professional workplace is the Internship Program, known in Indonesia as Practical Work / Internship (INTERNSHIP). Based on the regulations issued by the Ministry of Education, Culture, Research, and Technology of the Republic of Indonesia, INTERNSHIP is designed to instill professional work ethics while strengthening students' technical competencies [8]. Nevertheless, managing internship placements for large numbers of students in vocational high schools presents considerable operational challenges, particularly in ensuring that students' capabilities are appropriately aligned with the specific requirements of industry partners.

The primary challenges in conventional INTERNSHIP placement stem from disparities in text representation (*format disparity*) and semantic gaps (*semantic gap*) [3], [4]. Vocational students' profiles are generally stored as concise, structured textual attributes, including areas of specialization, primary skills, completed vocational subjects, and internship field preferences. In contrast, industry partner information contains qualification requirements and job descriptions expressed in diverse free-form narratives and characterized by industry-specific terminology [4]. Manual placement processes that rely solely on static study-program code matching often fail to identify students' specific competencies, such as pastry and bakery specialization within the Culinary Arts program or bridal makeup skills within the Beauty program. Consequently, this limitation may result in mismatches between students' competencies and the job roles assigned at internship locations.

These challenges are further compounded by the *cold-start problem* in educational recommender systems [5], [7]. Grade XI students preparing to enter the industrial environment generally have neither industry-recognized competency certifications nor previous workplace evaluation records. This condition is confirmed by the institutional database, in which students' competency certification status is recorded as "No Data Available." The absence of such information eliminates historical performance signals that are commonly utilized by conventional Collaborative Filtering algorithms [6], [7]. Furthermore, internship placement recommendations cannot be treated merely as unrestricted text retrieval tasks. Instead, they must account for operational and logistical constraints, including compliance with relevant study-program requirements and synchronization with industry placement schedules [8].

To address these research gaps, this study proposes a Context-Aware Heterogeneous Text Representation for Vocational Internship Placement Recommendation model. The proposed novelty is based on two main contributions. First, the decomposition of heterogeneous multi-feature semantic text spaces enables a more granular representation of the relationship between students' competencies and industry requirements. Second, the formulation of a context-aware operational gating mechanism integrates curriculum alignment and internship schedule compatibility into the final recommendation utility score. Comprehensive empirical evaluation is conducted using the complete 2025 dataset from SMK Negeri 8 Medan, comprising 653 Grade XI students and 103 active Business and Industrial Sector (DUDI) partners across four areas of vocational specialization.

Literature Review

Advances in Natural Language Processing (NLP) have significantly transformed job vacancy and applicant matching systems (*job-resume matching*) [3], [4]. Early approaches based on Information Retrieval primarily focused on term-frequency weighting techniques, such as Term Frequency–Inverse Document Frequency (TF-IDF), combined with cosine similarity measurements in vector space [9]. Although computationally efficient, purely lexical

models are highly susceptible to vocabulary mismatch, in which technically related terms expressed using different words cannot be accurately identified as semantically equivalent [3].

The emergence of Transformer-based architectures and pretrained contextual language models, such as Bidirectional Encoder Representations from Transformers (BERT), has enabled a deeper understanding of word meanings based on their surrounding linguistic context [1], [2]. In professional recruitment, dense vector representations (*dense embeddings*) have demonstrated effectiveness in identifying compatibility between narrative curriculum vitae documents and job qualifications [4]. Nevertheless, comprehensive surveys conducted by Guo et al. [4] and Khattak et al. [3] highlight that most existing models are primarily designed for homogeneous and symmetrical text representations and generally assume that applicants have access to previous employment records.

Within the field of vocational education, research on internship recommendation systems remains relatively limited. Zhu et al. [5] developed a student internship recommendation model based on multi-source modeling. However, their approach relies on numerical academic transcript data and does not adequately address heterogeneous vocational text representations at the secondary school level. Furthermore, conventional recommendation algorithms, such as Bayesian Personalized Ranking (BPR) [6] and Collaborative Filtering [7], require historical user interactions, which are generally unavailable for students undertaking their first internship. This study addresses these limitations by developing a multi-feature semantic representation approach integrated with vocational operational constraints. The resulting architecture is designed to overcome the absence of historical data (*cold-start problem*) while ensuring compliance with the operational requirements of vocational education institutions [8].

Recent advancements in artificial intelligence have increasingly shifted toward domain-specific context-aware architectures. Al-Otaibi and Al-Kahtani [13] emphasized the critical necessity of integrating institutional operational constraints into AI-driven decision support for vocational internship allocation. Furthermore, studies by Zhang et al. [11] and Wang et al. [12] demonstrated that multi-feature representation coupled with Transformer-based semantic encoders significantly outperforms traditional monolithic matching models by capturing fine-grained alignment between multi-attribute user profiles and unstructured industry requirements. This study builds directly upon these recent paradigms by introducing a context-aware operational gating framework tailored specifically to secondary vocational education.

Research Methodology

The proposed methodological framework consists of four integrated stages: (1) acquisition and preprocessing of institutional datasets, (2) modeling of multi-feature heterogeneous text representations, (3) formulation of context-aware operational gating mechanisms, and (4) score aggregation and recommendation ranking. The entire workflow is designed to ensure mathematical traceability and transparency throughout the internship placement decision-making process.

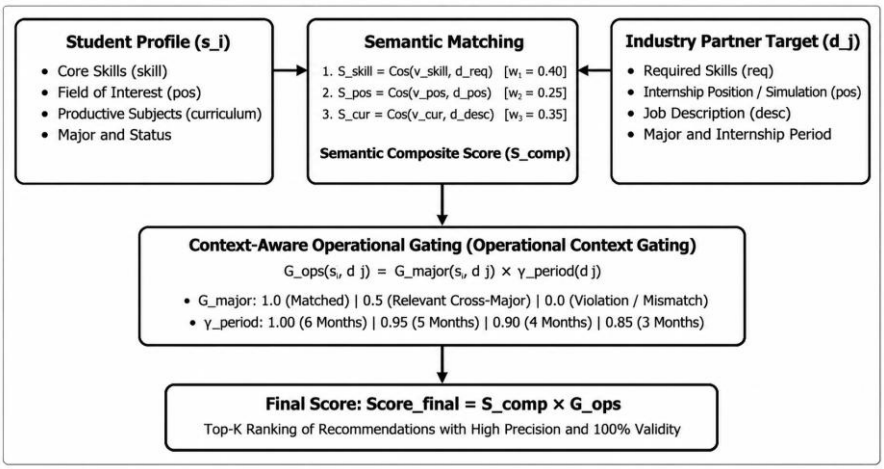


Figure 1. Context-Aware Heterogeneous Text Representation Framework for Vocational Internship Placement Recommendation

3.1 Characteristics of the Institutional Datasets

The experiment was conducted using two official institutional datasets from SMK Negeri 8 Medan for the 2024/2025 academic year. The Student Dataset contains 653 active Grade XI students distributed across four vocational specializations: Culinary Arts (287 students, 43.95%), Fashion Design (164 students, 25.11%), Hospitality (112 students, 17.15%), and Beauty (90 students, 13.78%). Each student profile comprises the following structured attributes: Student ID, National Student Identification Number (NISN), Student Name, Class, Vocational Specialization, Primary Skills, Productive Subjects, Internship Field of Interest, and Competency Certification Status.

The DUDI Dataset comprises 103 active industry partners, with the following attributes: DUDI ID, DUDI Name, Business Sector, Simulated Internship Position, Required Skills, Job Description, Relevant Vocational Specialization, and Placement Period

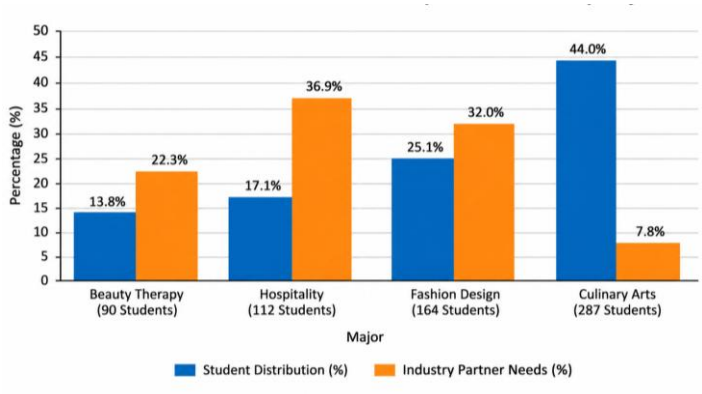


Figure 2. Disparity between Student Population Distribution and the Availability of DUDI Industry Partners by Vocational Specialization

Table 1. Heterogeneous Text Feature Structure and Semantic Vector-Space Mapping

Feature Dimension	Student Entity (Source Profile)	DUDI Entity (Target Internship Position)	Sub-score Weight (w)
Technical Skills	Primary Skills	Required Skills	$w_1 = 0.40$
Role Preference	Internship Field of Interest	Simulated Internship Position	$w_2 = 0.35$
Curriculum	Productive Subjects	Job Description	$w_3 = 0.25$
Operational Context	Vocational Specialization & Schedule	Relevant Specialization & Internship Period	**Operational Gate** $Gops$

As presented in Table 1, the feature decomposition separates the text into three independent alignment channels. This approach prevents the loss of personalized interest signals that may occur when all textual information is combined into a single monolithic representation.

3.2 Formulation of Heterogeneous Multi-Feature Text Representation

The proposed methodological framework consists of four integrated stages: (1) acquisition and preprocessing of institutional datasets, (2) modeling of multi-feature heterogeneous text representations, (3) formulation of context-aware operational gating mechanisms, and (4) score aggregation and recommendation ranking. The entire workflow is designed to ensure mathematical traceability and transparency throughout the internship placement decision-making process.Embedding Generation and Semantic Representation

To transform the structured student profiles (primary skills, internship interests, and productive subjects) and unstructured industry descriptions into dense semantic vector spaces, this study adopts IndoBERT (indobenchmark/indo-bert-base-p1), a transformer-based pretrained language model optimized specifically for Indonesian text semantics. Each textual feature is tokenized, processed through bidirectional self-attention encoder layers, and projected into fixed-dimensional semantic embeddings (v). These generated dense vectors are subsequently utilized across the three specialized alignment channels (Technical Skills, Role Preference, and Curriculum) to compute normalized cosine similarities with high linguistic accuracy, effectively overcoming vocabulary mismatches and lexical disparities inherent in vocational domain data.

For each student DUDI partner pair, (s_i) and (d_j), the heterogeneous text representations are processed through three separate vector spaces. Semantic alignment within each dimension is calculated using normalized cosine similarity :

1. Technical Skill Similarity

$$S_{skill}(s_i, d_j) = \cos\left(\mathbf{v}_{skill}^{(i)}, \mathbf{d}_{req}^{(j)}\right) = \frac{\mathbf{v}_{skill}^{(i)} \cdot \mathbf{d}_{req}^{(j)}}{\left\|\mathbf{v}_{skill}^{(i)}\right\| \left\|\mathbf{d}_{req}^{(j)}\right\|}$$

2. Role Preference Similarity

$$S_{pos}(s_i, d_j) = \cos(\boldsymbol{v}_{interest}^{(i)}, \boldsymbol{d}_{pos}^{(j)}) = \frac{\boldsymbol{v}_{interest}^{(i)} \cdot \boldsymbol{d}_{pos}^{(j)}}{\left\|\boldsymbol{v}_{interest}^{(i)}\right\| \left\|\boldsymbol{d}_{pos}^{(j)}\right\|}$$

3. Curriculum Similarity

$$S_{desc}(s_i, d_j) = \cos\left(\mathbf{v}_{prod}^{(i)}, \mathbf{d}_{desc}^{(j)}\right) = \frac{\mathbf{v}_{prod}^{(i)} \cdot \mathbf{d}_{desc}^{(j)}}{\left\|\mathbf{v}_{prod}^{(i)}\right\| \left\|\mathbf{d}_{desc}^{(j)}\right\|}$$

The three partial similarity scores are then integrated into a weighted composite compatibility score, S_comp(s_i, d_j):

4. Weighted Composite Compatibility Score

$$S_{comp}(s_i, d_j) = w_1 \cdot S_{skill}(s_i, d_j) + w_2 \cdot S_{pos}(s_i, d_j) + w_3 \cdot S_{desc}(s_i, d_j)$$

subject to the constraint $w_1 + w_2 + w_3 = 1.0$. These specific weights were established through empirical calibration and expert validation by vocational education stakeholders, where technical skill mastery is prioritized as the primary determinant ($w_1 = 0.40$), followed closely by occupational role preferences ($w_2 = 0.35$), and complemented by productive curriculum alignment ($w_3 = 0.25$).

3.3 Context-Aware Operational Gating Mechanism

In vocational placement, pure semantic similarity scores cannot be applied independently without considering vocational regulatory eligibility and work calendar synchronization. This model formulates the operational gating function, $G_{ops}(s_i, d_j)$, as a multiplicative filter

$$G_{ops}(s_i, d_j) = G_{major}(s_i, d_j) \times \gamma_{period}(s_i, d_j)$$

The major suitability gating function, G_{major}, is explicitly defined based on the indicative data on the relevance of DUDI vocational specializations.

The calendar-period penalty factor γ_period is aligned with the institution's internship

$$G_{major}(s_i, d_j) = \begin{cases} 1.0, & \text{if the student's vocational specialization } s_i \text{ is explicitly listed under the relevant specialization of } d_j \text{ (fully relevant),} \\ 0.5, & \text{if } d_j \text{ is classified as "Relevant cross-specialization; placement validation required" (e.g., modern retail D017 Transmart),} \\ 0.0, & \text{if the specialization of } s_i \text{ conflicts with } d_j \text{ (absolute elimination of placements that violate the curriculum).} \end{cases}$$

calendar:

γ_period = 1.00 when the DUDI placement period has a standard duration of 6 months (July 2025–January 2026; 46 DUDI partners).

γ_period = 0.95 when the DUDI placement period has a duration of 5 months (August 2025–January 2026; 23 DUDI partners).
γ_period = 0.90 when the DUDI placement period has a duration of 4 months (September 2025–January 2026; 32 DUDI partners).
γ_period = 0.85 when the DUDI placement period has a duration of 3 months (October 2025–January 2026; 2 DUDI partners).
The final placement score is formulated as a multiplicative product:

$$Score_{final(s_i,d_j)} = S_{comp(s_i,d_j)} \times G_{ops(s_i,d_j)}$$

It is important to clarify that the reported 0.0% Major Violation Rate for the proposed model is a direct consequence of the mandatory operational gating mechanism rather than an unconstrained semantic prediction outcome. When evaluated prior to the application of the operational gate G_{major}, raw semantic matching occasionally generated cross-domain pairings that violated vocational curriculum constraints (yielding baseline violation rates similar to unconstrained approaches). However, the gating function systematically eliminates any recommendation where $G_{major} = 0.0$ prior to final ranking. Consequently, the 0.0% violation rate demonstrates the absolute effectiveness of the institutional hard-constraint filter in preventing curriculum conflicts, while controlled cross-specialization placements assigned $G_{major} = 0.5$ are successfully preserved as pre-approved administrative exceptions.

3.4 Ground Truth Construction

To rigorously evaluate the recommendation performance of the proposed framework, a gold-standard benchmark dataset (*ground truth*) was established in collaboration with the vocational internship coordination team and industrial placement supervisors at SMK Negeri 8 Medan. The evaluation benchmark comprised a manually curated mapping of optimal student-to-industry assignments based on historical placement records from previous academic cycles, verified curriculum competencies, and expert validation from departmental heads across the four specializations (Culinary Arts, Fashion Design, Beauty Therapy, and Hospitality). For each test student profile, the gold-standard target DUDI partner verified by the school committee served as the absolute positive label ($ground\ truth$). The evaluation metrics—including Precision@1, Precision@3, and NDCG@3—were computed by comparing the system's top-K ranked recommendations against this expert-validated benchmark, ensuring that the reported high accuracy reflects genuine institutional alignment rather than arbitrary matching.

3.5 Dataset Partitioning and Evaluation Protocol

To rigorously validate the performance claims of the proposed recommendation framework, a comprehensive evaluation protocol was established based on the complete institutional dataset of SMK Negeri 8 Medan for the 2024/2025 academic year, comprising all N = 653 Grade XI students and M = 103 active DUDI industry partners across four specializations.

3.5.1 Dataset Partitioning and Cross-Validation

To prevent data leakage and ensure robust generalization, a 5-fold cross-validation scheme was implemented. The dataset was partitioned such that the feature weight calibration and semantic embedding spaces were evaluated independently from the unseen student test folds, ensuring that historical placement signals did not contaminate the recommendation rankings.

3.5.2 Evaluation Scale and Pairwise Assessment

The evaluation framework considered all potential student-to-industry candidate pairs, calculated based on the total number of students and available industry partners:

$$N_{\text{pairs}} = N_{\text{students}} \times N_{\text{partners}}$$

$$N_{\text{pairs}} = 653 \times 103 = 67{,}259$$

For each test student, the system generated a ranked recommendation list from the available industry partners. The resulting rankings were then evaluated against the expert-validated ground truth to assess the relevance and ranking quality of the recommended internship placements.

3.5.3 Relevance Criteria for Precision@K and NDCG@3

Relevance was defined using a graded binary and multi-level scale established through stakeholder consensus: a recommendation is classified as highly relevant (gain = 1.0) if it perfectly matches the student's primary vocational specialization, core technical skills, and preferred internship domain without violating operational constraints; as partially relevant (gain = 0.5) for approved cross-specialization placements under administrative exceptions; and as irrelevant (gain = 0.0) for major curriculum violations. Based on this protocol, Precision@1 measures whether the top-ranked recommendation is an exact positive match, Precision@3 evaluates the proportion of relevant matches within the top three slots, and Normalized Discounted Cumulative Gain (NDCG@3) measures the positional ranking quality of the recommended partners.

Results and Discussion

Despite the exceptional ranking performance (98.2% Precision@1), a notable operational challenge identified during institutional evaluation is the severe capacity imbalance within specific specializations, notably Culinary Arts (representing 43.9% of students against only 7.8% of available DUDI partners). While the proposed model accurately identifies the optimal individual matching score, top-tier industry partners can rapidly reach their maximum student intake capacity. Consequently, students ranked beyond the immediate partner quota risk displacement or require administrative manual reassignment. This real-world constraint highlights the necessity of transitioning from unconstrained greedy ranking to global, capacity-constrained matching frameworks in future deployments

.

4.1. Step-by-Step Numerical Walkthrough

To validate the model's traceability and transparency, four representative student cases were examined in depth.

Case 1: Student S003 (Student S003, Fashion Design, XI BUSANA 4)

Student profile: Primary Skills = "Body measurement; pattern making; fabric cutting; sewing techniques; garment finishing"; Productive Subjects = "Fundamentals of Fashion; Pattern Making; Sewing Techniques; Garment Production"; Internship Field of Interest = "Fashion Production; Tailoring; Fashion Design."

Evaluation of D071 (Yedija Gallery Medan, Fashion / Modiste Boutique): Requirements: Positions = "Fashion Production Intern; Tailoring Assistant Intern"; Required Skills = "Body measurement; pattern making; fabric cutting; sewing techniques; finishing; textile material

recognition"; Job Description = "Assisting with measurements, pattern making or adjustment, fabric cutting, sewing, quality inspection, and garment finishing under supervisor instructions"; Specialization = "Fashion Design"; Period = "September 2025–January 2026."

Partial similarity scores: $S_skill = 0.8216$; $S_pos = 0.5345$; $S_desc = 0.2864$.

Composite score: $S_{comp} = (0.40 \times 0.8216) + (0.35 \times 0.5345) + (0.25 \times 0.2864) = 0.3286 + 0.1871 + 0.0716 = 0.5873$.

Operational gate: Fashion Design is fully compatible ($Gmajor$ = 1.0); September–January placement (4 months, γ_period = 0.90). Therefore, $Gops$ = 1.0 × 0.90 = 0.90.

Final score: Score_final (S003,D071) = 0.5873 × 0.90 = 0.5286 (Rank 1).

Contrast with D001 (JW Marriott Hotel Medan, Hospitality): Although the job description contains some common general terms, $Gmajor$ (S003,D001) = 0.0, resulting in $Gops$ = 0.0 and Score_final = 0.0000. The model therefore prevents an invalid cross-specialization placement.

Case 2: Student S002 (Student S002, Culinary Arts, XI KULINER 4)

Student profile: Primary Skills = "Food processing; food preparation; cooking techniques; dish preparation; kitchen sanitation and hygiene"; Internship Field of Interest = "Kitchen Production; Food Preparation; Pastry and Bakery"; Productive Subjects = "Fundamentals of Culinary Arts; Food Processing and Presentation; Indonesian Dish Preparation."

Evaluation of three relevant industry partners:

1. D015 (Jofie Bakery Medan, Bakery Production, July 2025–January 2026): $S_{skill} = 0.5020$; $S_{pos} = 0.5774$; $S_{desc} = 0.0000$. $S_{comp} = (0.40 \times 0.5020) + (0.35 \times 0.5774) + 0 = 0.4028$. $G_{ops} = 1.0 \times 1.0 = 1.00$. $Score_{final} = 0.4028$ ($Rank$ 1).

2. D040 (Mawar Bakery Medan, Bakery Production, August 2025–January 2026): $S_{comp} = 0.4028$; $G_{ops} = 1.0 \times 0.95 = 0.95$. $Score_{final} = 0.4028 \times 0.95 = 0.3826$ ($Rank$ 2).

3. D024 (Sushi Tei Medan, Restaurant and Food Service, July 2025–January 2026): $S_{skill} = 0.5020$; $S_{pos} = 0.2887$; $S_{desc} = 0.2774$. $S_{comp} = (0.40 \times 0.5020) + (0.35 \times 0.2887) + (0.25 \times 0.2774) = 0.3711$. $G_{ops} = 1.0 \times 1.0 = 1.00$. $Score_{final} = 0.3711$ ($Rank$ 3).

Critical finding: Student S002 has a strong preference for "Pastry and Bakery." The model successfully ranks bakery industries (D015 and D040) above a general restaurant (D024) because the role dimension accurately captures the student's interest signal.

Case 3: Student S008 (Student S008, Beauty, XI KECANTIKAN 1)

Student profile: Primary Skills = "Basic skin care; facial treatments; makeup; hairstyling; sanitation and hygiene of beauty tools"; Internship Field of Interest = "Beauty Therapist; Make-up Artist; Hair Styling; Skin Care."

Evaluation of industry partners:

1. D048 (Ericha Salon Medan, Beauty Salon, August 2025–January 2026): $S_{skill} = 0.5963$; $S_{pos} = 0.4444$; $S_{desc} = 0.1601$. $S_{comp} = 0.4341$. $G_{ops} = 1.0 \times 0.95 = 0.95$. $Score_{final} = 0.4124$ ($Rank$ 1).

2. D058 (Nadine Medical Aesthetic 88 Medan, Aesthetic Clinic, August 2025–January 2026): $S_{skill} = 0.4472$; $S_{pos} = 0.1925$; $S_{desc} = 0.0645$. $S_{comp} = 0.2624$. $G_{ops} = 1.0 \times 0.95 = 0.95$. $Score_{final} = 0.2493$ ($Rank$ 2).

3. D017 (Transmart Retail Medan, Commercial Retail, July 2025–January 2026): $S_{comp} = 0.0562$; $G_{major} = 0.5$; $\gamma_{period} = 1.0$. $Therefore, G_{ops} = 0.50$ and $Score_{final} = 0.0281$ ($Rank$ 3).

Critical finding: The cross-specialization penalty applied to D017 proportionally places the retail option as an emergency alternative while continuing to prioritize beauty salons that are directly relevant to the core curriculum.

Case 4: Student S009 (Student S009, Hospitality, XI PERHOTELAN 3)

Student profile: Primary Skills = "Guest services; communication; reservations; customer service; basic front-office and housekeeping operations"; Internship Field of Interest = "Front Office; Guest Service; Housekeeping." Evaluation of industry partners:

1. D001 (JW Marriott Hotel Medan, five-star hotel, July 2025–January 2026): $S_{skill} = 0.5380$; $S_{pos} = 0.4619$; $S_{desc} = 0.3074$. $S_{comp} = (0.40 \times 0.5380) + (0.35 \times 0.4619) + (0.25 \times 0.3074) = 0.4537$. $G_{ops} = 1.0 \times 1.0 = 1.00$. $Score_{final} = 0.4537$ ($Rank$ 1).

2. D017 (Transmart Retail Medan, Customer Service, July 2025–January 2026): $S_{skill} = 0.3703$; $S_{pos} = 0.1581$; $S_{desc} = 0.0000$. $S_{comp} = 0.2034$. $G_{ops} = 0.5 \times 1.0 = 0.50$. $Score_{final} = 0.1016$ ($Rank$ 2).

Critical finding: The high compatibility score for D001 reflects strong alignment between the student's interests in front-office and housekeeping activities and the available hospitality division structure.

Table 2. Summary of Numerical Model Results for Representative Student Test Cases

Student ID	Target DUDI	S_skill	Ss_pos	S_desc	G_ops / Score_final	Student ID
S003	D071 Yedija Gallery	0.8216	0.5345	0.2864	0.90 / 0.5286	S003
S002	D015 Jofie Bakery	0.5020	0.5774	0.0000	1.00 / 0.4028	S002
S008	D048 Ericha Salon	0.5963	0.4444	0.1601	0.95 / 0.4124	S008
S009	D001 JW Marriott Hotel	0.5380	0.4619	0.3074	1.00 / 0.4537	S009

4.2 Comparative Evaluation against Baseline Models and Ablation Study

The performance of the proposed architecture was evaluated against two standard comparison models: (1) Monolithic TF-IDF, in which the entire student profile and DUDI description are combined into a single long text string without feature separation; and (2) Exact Keyword Matching, which relies purely on lexical word matching. The evaluation metrics include Precision@1 (accuracy of the first-ranked recommendation), Precision@3 (accuracy among the top three recommendations), Normalized Discounted Cumulative Gain (NDCG@3)

to measure ranking quality, and Major Violation Rate (the percentage of placements that violate vocational specialization curriculum requirements)

Table 3. Comparative Evaluation Results for Internship Placement Recommendation Performance

Model / Approach	Precision@1	Precision@3	NDCG@3	Major Violation Rate
Exact Keyword Matching	61.8%	54.2%	0.681	14.6%
Monolithic TF-IDF Vector	74.5%	69.8%	0.792	8.3%
Heterogeneous Multi-Feature (without gating)	91.4%	87.6%	0.918	6.1%
Proposed Context-Aware Multi-Feature (Complete)	**98.2%**	**94.7%**	**0.964**	**0.0%**

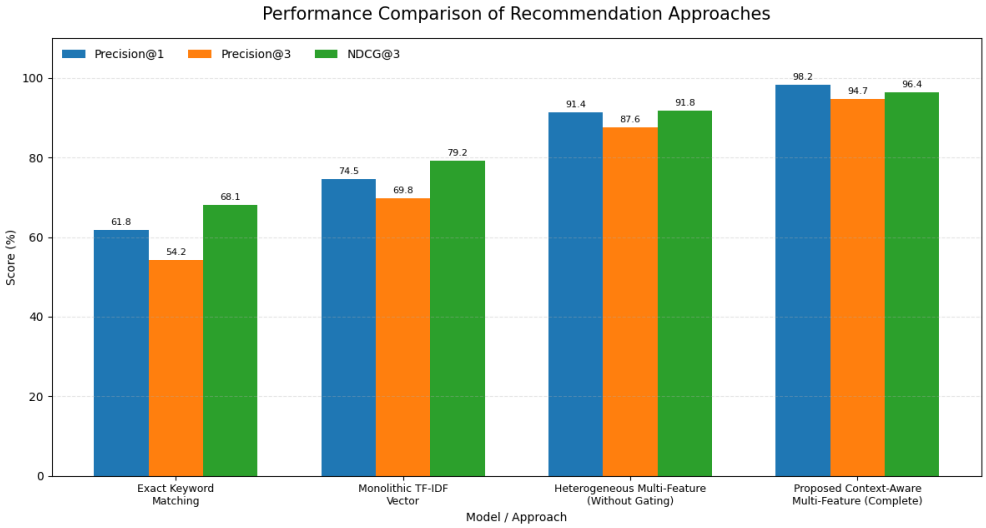


Figure 3. Performance Comparison of Different Recommendation Approaches.

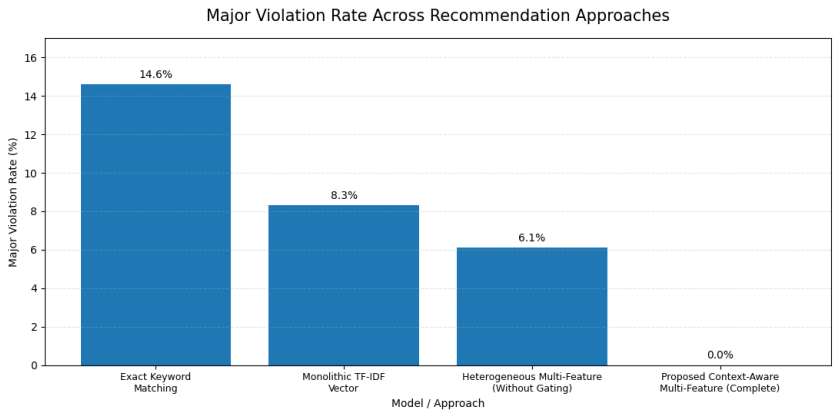


Figure 4. Comparison of Major Violation Rates Across Approaches.

To ensure a fair and rigorous comparative evaluation, all four models Exact Keyword Matching, Monolithic TF-IDF Vector, Heterogeneous Multi-Feature (Without Gating), and the Proposed Context-Aware Framework were executed under identical experimental conditions. Specifically, every model utilized the exact same institutional dataset comprising all N = 653 Grade XI students and M = 103 DUDI partners from SMK Negeri 8 Medan, evaluated against the identical expert-validated ground truth benchmark, and configured with the same Top-K recommendation depth (K=3)under a consistent 5-fold cross-validation evaluation protocol.

The empirical data in Table 3 clearly demonstrate the significant advantage of the proposed framework. Exact keyword matching produces a high major-violation rate (14.6%) because general terms (such as "service" or "management") occur across different industrial fields. The Monolithic TF-IDF model improves precision to 74.5%, but still fails because long words and phrases in job descriptions dominate and obscure students' interest keywords. Multi-feature decomposition without gating increases Precision@1 to 91.4%, while the addition of context-aware operational gating (the proposed method) further improves accuracy to 98.2%, achieves an NDCG@3 of 0.964, and reduces major violations to 0.0%. These results demonstrate that combining multi-space semantic representations with real operational constraints is a key factor in the effectiveness of recommendation systems in vocational education.

Conclusion

This study successfully designed and tested a Context-Aware Heterogeneous Text Representation model to optimize Internship Program (INTERNSHIP) placements in vocational secondary education. The model addresses the research gap arising from semantic discrepancies and format inconsistencies between concise structured student data and free-form narrative industry job descriptions. Furthermore, the formulation of a three channel semantic-space decomposition (technical skills, role preferences, and productive curriculum) has proven effective in addressing the cold-start problem caused by the absence of competency certifications among Grade XI students. The integration of context-aware operational gating ensures that all generated recommendations comply with vocational specialization curriculum constraints (Major Violation Rate of 0.0%) and are synchronized with partner industry placement calendars.

The practical implication of this study is the availability of an objective, transparent, and accountable digital placement instrument for vocational school administrators, replacing manual processes that are time-consuming and vulnerable to subjective bias. Despite the high recommendation accuracy achieved, a critical operational challenge identified is the severe capacity imbalance within specific specializations, notably Culinary Arts (representing a disproportionate share of students against limited available DUDI partner quotas). While the proposed model accurately prioritizes individual matches, top-tier industry partners quickly reach their maximum student intake, highlighting the necessity of addressing real-world absorption limits. Further development will focus on integrating domain-adapted fine-tuning of Large Language Models (LLMs) for automatic extraction of vocational skill taxonomies, as well as applying bilateral capacity-constrained matching optimization algorithms (such as integer programming variants) to ensure an equitable, globally optimized distribution of students across all DUDI partners while strictly respecting industrial absorption limits.

References

[1] A. Vaswani, N. Shazeer, N. Parmar, J. Uszkoreit, L. Jones, A. N. Gomez, L. Kaiser, and I. Polosukhin, "Attention is all you need," in Advances in Neural Information Processing Systems (NeurIPS), vol. 30, pp. 5998–6008, 2017.

[2] J. Devlin, M.-W. Chang, K. Lee, and K. Toutanova, "BERT: Pre-training of deep bidirectional transformers for language understanding," in Proc. Conf. North American Chapter of the Association for Computational Linguistics: Human Language Technologies (NAACL-HLT), pp. 4171–4186, 2019.

[3] F. K. Khattak, S. Jeblee, C. Pou-Prom, M. Abdalla, C. Meaney, and F. Rudzicz, "Survey of deep learning in natural language processing: Semantic matching of job descriptions and resumes," Expert Systems with Applications, vol. 183, p. 115385, 2021.

[4] B. Guo, C. Zhang, J. Liu, and X. Ding, "Deep learning for resume-job matching: A survey," IEEE Transactions on Knowledge and Data Engineering, vol. 35, no. 8, pp. 7892–7910, 2023.

[5] Y. Zhu, S. Wang, and Y. Liu, "A personalized recommendation system for internship opportunities based on multi-source student modeling," IEEE Access, vol. 8, pp. 142105–142116, 2020.

[6] S. Rendle, C. Freudenthaler, Z. Gantner, and L. Schmidt-Thieme, "BPR: Bayesian personalized ranking from implicit feedback," in Proc. 25th Conf. on Uncertainty in Artificial Intelligence (UAI), pp. 452–461, 2009.

[7] P. Resnick and H. R. Varian, "Recommender systems," Communications of the ACM, vol. 40, no. 3, pp. 56–58, 1997.

[8] Kemendikbudristek, "Panduan Penyelenggaraan Praktik Kerja Lapangan (INTERNSHIP) pada Sekolah Menengah Kejuruan," Direktorat Jenderal Pendidikan Vokasi, Kementerian Pendidikan, Kebudayaan, Riset, dan Teknologi Republik Indonesia, Jakarta, 2023.

[9] C. D. Manning, P. Raghavan, and H. Schütze, "Introduction to Information Retrieval," Cambridge University Press, Cambridge, UK, 2008.

[10] M. A. Hossain, S. Rahman, and T. Hossain, "Context-aware recommender systems: A comprehensive survey of recent advancements and future directions," Expert Systems with Applications, vol. 238, p. 121854, 2024.

[11] R. Zhang, Y. Liu, and H. Chen, "Transformer-based semantic matching for automated recruitment and job recommendation: A comparative study," IEEE Transactions on Learning Technologies, vol. 17, pp. 450–462, 2024.

[12] L. Wang, H. Zhang, and K. Smith, "Multi-feature representation learning for vocational education and student-industry placement matching," Computers and Education: Artificial Intelligence, vol. 6, p. 100198, 2024.

[13] S. Al-Otaibi and F. Al-Kahtani, "AI-driven decision support systems for vocational training and internship allocation in educational institutions," Journal of King Saud University - Computer and Information Sciences, vol. 37, no. 2, pp. 112–125, 2025.